鲁棒可预测判别字典学习人脸识别方法 *

张　健 [a, b]，米建勋 [a, b†]

(重庆邮电大学 a. 计算机科学与技术学院; b. 计算智能重庆市重点实验室，重庆 400065)

摘　要：提出一种可预测判别 K-SVD 网络模型(DKSVDN)并用于人脸识别问题。该模型构造了一种新颖的字典结构，包含类别标签字典和描述字典，以兼顾判别和重构性能。相应的稀疏编码向量由标签编码向量和描述编码向量组成。针对样本稀疏编码时间效率低的问题，利用预测神经网络与判别字典学习模型协同训练的方法来加速预测稀疏编码。此外，针对 DKSVDN 还特别引入一种拟梦境的训练方法用于提升模型在训练集多样性不足时的鲁棒性。通过在主流人脸数据集上的对比实验证明了该模型的优良性能。

关键词：字典学习；稀疏表示；人脸识别；神经网络

中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2020.03.0080

Robust discriminative K-SVD network for face recognition

Zhang Jian[a, b], Mi Jianxun[a, b†]

(*a. School of Computer Science & Technology*, *b. Chongqing Key Laboratory of Computational Intelligence*, *Chongqing University of Posts & Telecommunications*, *Chongqing* 400065, *China*)

Abstract: This paper presented a novel discriminative K-SVD network (DKSVDN) for face recognition. It embedded discriminative information into traditional K-SVD algorithm by special design of dictionary as well as sparse representation coefficients on the dictionary. The dictionary consisted of label specific atoms and descriptive atoms, while sparse codes contained one-hot label vectors and descriptive codes. In addition, as sparse representation algorithms were time-consuming, DKSVDN attached a co-trained feed-forward neural network to discriminative dictionary learning model to predict sparse codes. Moreover, with generative module in DKSVDN, this work also designed a new dreaming training phase to improve the robustness of DKSVDN for unknown pattern in known class. The experiment results on public face image datasets verified effectiveness of this method.

Key words: dictionary learning; sparse representation; face recognition; neural network

0　引言

近年来，针对基于生物特征识别的身份识别技术的需求日益增长。人脸识别是生物特征识别的重要技术，得益于其无须接触、识别快速以及方便部署等特点，获得了社会各界的广泛关注，并迅速成为了计算机视觉领域最为热门的研究方向之一。人脸识别技术[1-3]最初基于几何特征对比，随后引入数据驱动的算法模型，包括以主成分分析、特征脸为代表的基于机器学习的人脸识别方法[4-11]以及基于残差神经网络、谷歌网络等深度神经网络的深度人脸识别方法[12-17]，这些方法的识别精度不断提升。然而，由于现实环境中存在大量的不确定性因素，人脸识别算法在实际应用于现实场景时需要面对的诸多问题更为复杂，光照、随机噪声、表情、姿态和装饰遮挡往往不能在训练数据中得到充分考虑。对现实场景的不充分采样严重影响了数据驱动的人脸识别算法模型的识别效果，成为了当下必须解决的问题。

字典学习(dictionary learning)是一种基于压缩感知理论(compressed sensing theory)[18]的生成模型，由 Olshausen 与 Field[19]于 1996 年首次提出。此前，在生物学领域中关于哺乳动物的视觉机制的研究有了重大突破，研究发现，在视觉通路上的很多神经元对于初级视觉和中级视觉中的特殊刺激具有选择性，例如，颜色、纹理、方向、尺寸，甚至是不同视图的物体图像。基于这样的事实，Olshausen 与 Field 设计了一种基于能量的字典学习模型。通过对样本图像训练学习，字典学习模型表明在过完备字典上自然图像具有十分紧凑的稀疏表达(sparse representation)系数，即图像可以由过完备字典中少量的字典项通过线性组合重构，而这很可能就是大脑视觉皮层 V1 区域所采取的编码策略。同时，在文献[20]中提出了一种二元交替优化结构来求解模型，优化过程具体包括稀疏编码部分和字典构造部分。这种结构奠定了字典学习的基础，被后继的字典学习算法普遍采用。Aharon 等人[21]通过探索向量量化与字典学习算法的联系提出了著名的 K-SVD 方法，该方法通过推广 K-means 算法改进了基于能量的字典学习模型。K-SVD 算法开创性地在字典构造阶段引入了奇异值分解方法(SVD，singular value decomposition)来最小化重构误差。相较于已有字典学习算法存在只能批处理优化的局限，在线字典学习(ODL，online dictionary learning)[22]提出了在线训练字典的方法，在现如今海量的训练图片更具优势。ODL 利用 L_1 范数约束模型的稀疏性，在字典构造阶段采取二阶优化求解最优字典项，这使得随机优化和小批量样本优化得以实施。字典学习的算法框架是在图像重构领域获得了巨大的成功[23]，许多的研究者将目光投向了热门的图像分类任务。Mairal 等人[24]提出了任务驱动的字典学习算法(task-driven dictionary learning)，将监督信息引入了传统字典学习，将字典学习算法用于手写数字识别。D-KSVD (discriminative K-SVD)[5]扩展了 K-SVD 算法，将字典学习算法用于人脸识别

收稿日期：2020-03-07；**修回日期：**2020-04-26　**基金项目：**重庆市自然科学科学基金资助项目(cstc2018jcyjAX0532)；国家自然科学基金资助项目(61906024)

作者简介：张健(1994-)，男，江苏泰兴人，硕士研究生，主要研究方向为人工智能与模式识别；米建勋(1982-)，男(通信作者)，重庆人，博士，主要研究方向为人工智能、模式识别、智能计算(mijianxun@gmail.com).

任务。D-KSVD 利用 K-SVD 训练字典学习的同时训练一个专用分类器用于图像分类的任务，其在多个人脸图像数据集上都展现了优异的性能。Jiang 等人[25]基于 DKSVD 提出了一个改进算法 LCKSVD(Label Consistent K-SVD)。LCKSVD 在 DKSVD 模型中加入了一个标签回归项以获得更强的判别信息。相较于上述字典学习分类方法使得稀疏编码携带分类信息，还有一类方法则构造具有判别性的字典，其基本思想是：针对每个类构造该类的字典，而样本在每个类别字典上重构误差往往被用来作为分类依据，即样本属于重构误差最小的子字典所属的类[10, 11, 26]。值得一提的是，这种方法是建立在样本和同一类的样本更加相似这个先验认知上的。这一类方法的起源是稀疏表达分类方法(SRC，sparse representation based classification)[10]。SRC 直接利用所有类别训练样本构造判别性字典来编码新样本。它在分类任务中获得令人惊喜的表现，但 SRC 的缺陷也是显而易见的：为提升最终分类性能所需要的字典规模往往过于庞大。随后，DLSI(structural incoherent dictionary learning method)[27]针对 SRC 作出了改进，使得字典规模更为紧凑。以上提及的两类监督字典学习方法分别从两个不同的方向提升字典学习模型的判别性能。此外，研究者们还提出了同时使得编码向量和字典兼具判别性的方法。比如，FDDL(Fisher discrimination dictionary learning)[28]就是一种典型的编码向量和字典兼具判别性的方法。FDDL 应用 Fisher 判别准则学习一个结构化的字典，与此同时，Fisher 判别准则还被应用于编码系数上使得编码向量具有尽可能大的类间距离和尽可能小的类内距离，从而获得优异的判别性能。Wang 等人[29]在中提出了一种同时学习类别字典和通用字典的模型(DLSPC)。这种方法利用类内字典抓取每一类的最具判别性的细节特性而公共字典用于保存共享的元素。此外，DLSPC 还对编码加以约束，使得样本只在同一类的字典上进行表达而抑制在不同类字典上的表达。DLSPC 利用样本在每一类子字典上重构误差进行分类。然而由于字典学习模型依赖于稀疏编码，在实际应用中需要额外的计算来获取编码系数，这使得基于字典学习的模型难以大规模应用。LeCun 等人[30]尝试使用神经网络来预测稀疏编码系数，获得了较好的效果。Wang 等人[31]提出深度 L_0 编码器，利用深度网络结构来模拟 L_0 范数求解，获得稀疏编码。

本文提出一种可预测判别 K-SVD 网络，通过设计一种新颖的字典结构将标签信息嵌入 K-SVD 算法中，使得稀疏编码直接包含类别标签，相较于其他字典学习分类，无须额外计算即可直接得出类别；并且该字典结构将细节描述性信息剥离与本质特征分开存放，字典的解释性更强，并能用于生成样本。同时，训练编码预测网络模型，提升计算稀疏编码的效率。此外，本文引入一种拟梦境训练方式，通过超采样训练数据集生成虚拟样本用于预测网络的训练，提高预测准确性。在主流人脸图像数据库中的实验效果证明本文提出的方法在人脸图像识别中具有良好的表现。

1　相关工作

$D=[d_1,d_2,\cdots,d_k]\in\mathbb{R}^{m\times k}$ 表示字典矩阵，其每一列为一个字典元素向量。在经典的稀疏表达问题中，一个样本向量 $x\in\mathbb{R}^m$ 可以被表示成这些原型元素的线性组合:

$$x = D\alpha = \alpha^1 d_1 + \alpha^2 d_2 + \cdots + \alpha^k d_k \tag{1}$$

其中，$\alpha=[\alpha^1,\alpha^2,\cdots,\alpha^k]^T$ 是 x 在字典 D 上的稀疏表达系数向量。由于字典 D 是过完备的，稀疏表达问题往往有无穷多个解，为了获得本文所需要的最终解，这就需要对对原始问题设置合适的稀疏约束条件。显然，稀疏表达的解应为如下问题的解：

$$\min_{\alpha}\|\alpha_0\|\quad s.t.\ x = D\alpha \tag{2}$$

或者

$$\min_{\alpha}\|\alpha\|_0\quad s.t.\ \|x-D\alpha\|_2\le\epsilon \tag{3}$$

其中，$\|*\|_0$ 为 L_0 范数，表示向量 α 中的非零元素的个数。

稀疏表达解决了在特定字典上的最优编码问题，然而，优秀的稀疏编码还要求一个合适的字典。预先设计好固定字典是一种常用的方法，但是设计一个好的预设字典往往需要耗费大量的时间与精力，并且随机因素对结果的影响较大。此外，为了尽可能地涵盖所有需要的特征，预设字典一般会设置为过冗余。为了克服预设字典的各种缺陷，字典学习算法被提出来用于从训练数据集中学习最合适的字典。

在字典学习问题中，除了基本的稀疏表达求解，还需要从训练集中学习一个用于编码的最优的字典矩阵 D 。记训练样本集为 $X=[x_1,x_2,\cdots,x_n]\in\mathbb{R}^{m\times n}$ ，则字典学习问题可以被表达为

$$\min_{\langle D,A\rangle}\|X-DA\|_F^2\quad s.t.\ \forall i,\ \|\alpha_i\|_0\le T_0 \tag{4}$$

或者也可以写为

$$\min_{\langle D,A\rangle}\sum_i\|\alpha_i\|_0\quad s.t.\ \|X-DA\|_F^2\le\epsilon \tag{5}$$

其中，ϵ 为固定阈值。$A=[\alpha_1,\alpha_2,\cdots,\alpha_n]\in\mathbb{R}^{k\times n}$ 是 X 在字典 D 上的稀疏编码矩阵。$\|*\|_F$ 为 Frobenius 范数，用于计算重构误差。本文中主要考虑式(5)。同样，由于含有 L_0 范数的问题求解是非凸优化问题，式(5)不存在闭式解。在基于 L_0 范数的字典学习问题中广泛采用交替迭代的优化策略来最小化表达式(5)：首先固定字典 D ，计算基于当前字典的最优稀疏表达系数矩阵 A ；其次，固定稀疏系数矩阵 A ，优化字典 D 使得重构误差最小。交替迭代两个步骤，直至收敛。

训练过程完成后，最终获得的字典 D 在测试过程中不再变化，用于编码新样本，而新来样本的编码的求解则变为求解式(2)(3)。

K-SVD 作为最著名的字典学习方法之一，获得了广泛的关注。KSVD 算法中在稀疏编码阶段利用匹配算法或正交匹配算法求解系数向量。在字典更新阶段，式(5)中的矩阵相乘项 DA 被分解为 K 个秩为 1 的矩阵。更新过程中，每次只更新其中一个矩阵而固定剩下的 $K-1$ 个矩阵，如此逐个优化，最后使得重构误差最小。因此，重构误差项可以被改写为

$$\|X-DA\|_F^2=\left\|X-\sum_{j=1}^{K}d_j\alpha_T^j\right\|_F^2=\left\|\left(X-\sum_{j\ne k}d_j\alpha_T^j\right)-d_k\alpha_T^k\right\|_F^2=\|E_k-d_k\alpha_T^k\|_F^2 \tag{6}$$

其中，α_T^k 代表系数矩阵 A 的第 k 行。E_k 表示去除第 k 个字典项后，由剩下的字典项表达样本时所产生的重构误差。利用奇异值分解 SVD 分解误差即 $E_k=U\Delta V^T$ ，取 U 第一主分量即第一列作为字典项 d_k ，对应地，奇异值矩阵 Δ 第一行为所求 α_T^k ，但由于需要保证稀疏表达系数地稀疏性，KSVD 中使用了一个小技巧，最终字典项 d_k 稀疏系数 $\alpha_R^k=\Delta(1,1)\times V$ ， $\Delta(1,1)$ 为 Δ 的第一行第一列的值。K-SVD 在不同的测试数据集以及信号与图像重构任务上都表现出了较好的性能。

2　本文算法

K-SVD 是基于 L_0 约束的字典学习算法，算法框架包含稀疏编码更新和字典更新两个子步骤。在优化过程中，两个步骤交替迭代，直至收敛。通过观察可以发现，整个优化过程隐含了一个自动编码机，其中稀疏编码子步骤为一个非参的编码器而字典重构的部分则为线性解码器。

本文提出了一种新颖的判别字典学习模型，并同时构造了一个前向神经网络预测判别稀疏编码系数，以提高算法时间效率。

2.1　算法描述

本文提出判别 K-SVD 网络模型 DKSVDN(discriminative K-SVD network)，在原始 KSVD 基础上加入判别信息，并提高稀疏编码的计算效率，其目标函数表示为

$$\min_{A,D,\mathcal{P}_f} R(X,D,A) \quad \text{s.t.}\ \forall i, \|\alpha_i\|_0 \le T_0, P(\mathcal{P}_f; X, A) \le \epsilon \tag{7}$$

其中，$R(X,D,A)$ 是判别字典学习模块，$\|\alpha\|_0$ 是 L_0 稀疏约束项，每个样本对应的稀疏系数向量中取值非零的维度应小于 T_0，$P(\mathcal{P}_f; X, A)$ 则是预测神经网络模块，X 为输入样本矩阵，A 为目标稀疏矩阵，$\mathcal{P}_f$ 为前向神经网络的所有参数的集合。

2.1.1 判别字典学习模块

稀疏表达系数具有天然的优良特征，其能很好在图像重构和分类任务中都表现出了很好的鲁棒性。然而，重构性能和判别性能在传统字典学习模型中往往很难兼顾，本文设计了一种新颖的稀疏编码与字典结构，同时保持两种特性。该模型将类别标签嵌入稀疏编码中，使得编码包含标签编码和描述编码两部分，相对应地，对应的字典也同时分为了标签字典和描述字典两个子字典。假设 $Y=[y_1,y_2,\cdots,y_n]$ 表示标签矩阵，其中每一列 $y_i(i=1,2,\cdots,n)$ 代表训练样本 x_i 的标签向量，$y_i=[y_i^1,y_i^2,\cdots,y_i^c]^T$ 为 $0-1$ 向量，其中 c 为类别数，若 x_i 属于第 s 类则 y_i 的第 s 维 y_i^s 取值为 1，其他维度取值为 0。那么，稀疏编码向量具体表示为 $\alpha=\begin{bmatrix} y \\ \alpha_d \end{bmatrix}$，其中 y 为 0-1 标签向量，α_d 为描述编码。标签向量 y 对应于标签子字典 D_l，D_l 包含的字典项数和类别数相同，α_d 对应的描述字典为 D_d。标签字典 D_l 用于捕获每一类中最为本质的特征，这些特征一般代表了类别的特性，位于类别中心；描述字典 D_l 则用于保存无关类别信息的特征，这些特征可能出现于任何类别的通用特征，用于描述图像中的细节信息。因此，字典具体表示为 $D=[D_l,D_d]$，判别字典学习模块定义为

$$R(X,D,A)=\left\|X-[D_l,D_d]\begin{bmatrix} y \\ \alpha_d \end{bmatrix}\right\|_F^2 \tag{8}$$

2.1.2 预测神经网络模块

给定输入样本图像 x，预测神经网络模块用于快速预测样本的稀疏系数。相较于传统稀疏编码方法利用迭代优化求解稀疏编码而在测试阶段需要花费较长时间计算时间且时间长度无法预测，预测神经网络模块使用参数化的模型：神经网络。结合非线性激活函数，神经网络具有强大的近似拟合能力，并能有效大幅减少计算时间，在可预期的时间内给出较为准确的稀疏表达系数，以便后续任务使用。本文中使用多层前向全连接网络 $F(\mathcal{P}_f, X)$ 作为预测模块模型，为了获得稀疏性，激活函数采用软阈值函数，因此，预测神经网络模块的表达式为

$$P(\mathcal{P}_f; X, A)=\|F(\mathcal{P}_f, X)-A\|_F^2 \tag{9}$$

综上所述，最终目标函数为

$$\min_{A,D,\mathcal{P}_f} \left\|X-[D_l,D_d]\begin{bmatrix} y \\ \alpha_d \end{bmatrix}\right\|_F^2 \quad \text{s.t.}\ \forall i, \|\alpha_i\|_0 \le T_0, \|F(\mathcal{P}_f, X)-A\|_F^2 \le \epsilon \tag{10}$$

其中，ϵ 为固定阈值。利用稀疏编码分类时，首先提取稀疏编码中标签编码部分，标签编码中各维度中表达值最大的那个维度即为测试图片的预测类别。值得一提的是，为了更好的预测效果以及避免过拟合，本文采取协同训练的方式训练预测神经网络模块和判别字典学习模块。

2.2 算法优化求解

目标函数显然是非凸的，本文利用交替优化的策略求解该模型，具体优化流程如下：

a) 固定字典 D，利用正交匹配追踪(Orthogonal Matching Pursuit, OMP)[32]搜索最优在当前字典上的最优稀疏表达系数。更具体地，标签向量 y 为已知监督信息固定不变，优化目标为最优描述系数部分 u_d；

b) 将步骤 a) 获得的稀疏表达系数作为已知固定值，利用 K-SVD 算法逐个更新字典项，包括标签字典 D_l 和描述字典 D_d。同时，将稀疏稀疏矩阵作为目标值，训练样本矩阵作为输入，利用随机梯度下降(stochastic gradient descent,SGD)更新预测神经网络模块的参数集合 $\mathcal{P}_f$，直至收敛或者到达固定迭代步数。

交替迭代步骤 a)b)直至收敛或者达到固定步数则结束。另外，预测神经网络模块地预测值可以作为优化稀疏表达的初始值作为热启动值(warm start)以加快步骤 a)的收敛速度。

2.3 拟梦境训练法

已知类别未知模式(unknown pattern in known class,UPKC)是现实环境中人脸识别所面对的非常广泛也非常关键的挑战之一。UPKC 包含多种人脸变化，包含但不限于表情变化，遮挡变化等等。然而，不同于广泛的人脸变化中包含大量的随机信息，UPKC 更关注于具有规律性的人脸变化模式，对于人脑来说这些变化往往可以预测与想象，可以从一个类别推广至所有类别。具体举例来说，假设某一类中存在一种独有模式：人脸上戴有某一花纹的围巾，人类大脑可以轻易通过该张图片与其他人脸图片想象出其他类别中戴有相同围巾的人脸图片，但是对于统计机器学习算法，由于其他类别均缺失该模式，该围巾反而会被学习作为类别特征，从而出现错误分类。类似的模式有配戴墨镜，不同程度地笑容等等。由于采集所有人的人脸图像的所有模式显然是不可能的任务，因此，UPKC 是必须面对的与解决的问题。

为了提高 DKSVDN 针对 UPKC 的鲁棒性，本文设计了新颖的拟梦境训练方法。拟梦境训练方法的思想十分类似于人类做梦的过程与作用。文献[33]中首次提出了相似的思想，主要用于训练亥姆霍兹机(Helmholtz machine)。之后，同样的方法亦被用于训练深度信念网络(Deep Belief Networks, DBN)[34]。如前所述，DKSVDN 模型中包含了特殊结构的字典，该字典由类别字典和标签字典组合构成。相对应地，样本在该字典上的稀疏编码分为类别标签向量段和描述向量段。拟梦境训练方法充分利用了 DKSVDN 的结构特点，在稀疏编码空间合理推断 UKPC 的稀疏编码向量，借助字典生成虚拟训练样本以训练预测神经网络模块，提高预测准确度。

拟梦境训练方法结合 DKSVDN 具体步骤如下：a) 收集记录样本在当前字典上的稀疏编码；b) 拆分稀疏编码中的描述编码段，整理收集至描述编码池中；c) 对于所有类别编码向量，随机选取描述编码池中的描述编码向量进行拼接组合，生成合理推断稀疏编码向量；d) 利用生成的合理推断稀疏编码向量结合字典，通过式(1)生成虚拟样本；e) 将虚拟样本和对应的合理推断稀疏编码向量作为训练集，训练预测神经网络模块。

拟梦境训练利用生成字典学习模型在稀疏域进行样本合成。相较于已有的过采样方法，拟梦境训练充分考虑了样本的合理性，使得虚拟样本空间分布更加贴近现实样本空间，有效提高预测神经网络的预测准确度。拟梦境训练可以插入 DKSVDN 模型优化过程中任意一次迭代之后。

3 实验分析

3.1 实验设置

3.1.1 数据集介绍

为了验证 DKSVDN 在不同人脸识别场景中的性能，本文在主流的人脸数据库：AR 人脸数据库[35]与 Extend Yale B 人脸数据库[8]上进行实验。

AR 人脸数据库采集了 126 个人在不同条件下的人脸图片样本，总共包含超过 4000 张图片。该数据集内分为了两个子集，两个子集在不同时间段拍摄。在 AR 数据集中，每个人，即每个类别，拥有 26 张在不同的光照条件下拍摄面部照片，其中每个人有 12 张图片带有遮挡，具体为太阳镜遮挡和围巾遮挡，其他 14 张图片无遮挡但拥有表情与姿态变化。相

较于 Extend Yale B 数据集，AR 数据集包含了除光照以外的更多的干扰因素，包括表情与姿态的变化，面部遮挡等，这些都使得 AR 数据集更加贴近现实环境，也更具有挑战性。在本文实验中，使用了由该数据集中 120 人的 3120 张图像组成的 AR 人脸数据库的子集。其中，每个类别任意选取 13 张图片作为训练数据，剩下的 13 张组成测试集。所有图像的分辨率统一剪裁为 40×50 像素，并进行 L_2 范数归一化。图 1 展示了 AR 数据集中的部分人脸图像。



图 1　AR 人脸数据库中人脸图像

Fig. 1　Face images from AR face database

Extend Yale B 人脸数据库由 2414 张正面人脸图像组成，包含了 38 个人的人脸图像，这些图像在不同的光照条件和表情下拍摄。除光照外，这些人脸图像的变化还体现在表情变化上。其中，每个人大约有 59 到 64 张图像，为了方便处理，本文中选取 Extend Yale B 人脸数据库中的 31 个人，每个人均拥有 64 人脸图片，总共 1984 张图片，去除了 7 个人，属于这些人图片均不足 64 张。其中，每个类别任意选取 32 张图片作为训练数据，剩下的 32 张组成测试集。所有图像的分辨率统一剪裁为 40×50 像素，并进行 范数归一化。图 2 展示了 Extend Yale B 数据集中的部分人脸图像。



图 2　Extend Yale B 人脸数据库中人脸图像

Fig. 2　Face images from Extend Yale B face database

3.1.2 参数设置

在预测编码模块中，利用前向神经网络作为预测模块，该神经网络具有两层结构，两层神经元的数目分别是 1500 与 500。字典项的数目为 500，其中标签字典项的数目对应于 AR 人脸数据库与 Extended YaleB 人脸数据库分别为 100 项和 31 项。T_0 值设置为 15。

3.1.3 分类模式

本文所提出的 DKSVDN 模型不需要训练额外的分类器，具体类别信息可以直接从稀疏编码中的标签向量部分读取。由于预测神经网络模块不能获得真正完美的 0-1 标签向量，因此测试样本的类别判定为表达值最大的维度对应的类别，数学表达式可记为

$$c = \max_s \hat{y}^s \tag{11}$$

其中，$\hat{y}$ 为预测标签向量，$\hat{y}^s$ 表示 $\hat{y}$ 第 s 维的取值。利用训练好的模型，稀疏编码向量有两种获取方式。首先，可以利用预测神经网络模块直接预测稀疏编码向量。为了获得更加精确的稀疏编码，依然可以将预测稀疏编码向量作为起始值，通过 KSVD 算法优化稀疏编码向量。这两种计算方式在后续实验同时进行测试。

3.2　实验结果

3.2.1 人脸识别

本文中参与对比的算法有 DLSI，LCKSVD，FDDL，FDDL-LCSRC[36]，N-PCA[37]，RCSRC[38]。在 AR 人脸数据集上，DLSI 每个类别的类内字典项设置为 5，其他设置于文献[25]中一致。根据 AR 数据库中的类别数，LCKSVD 字典数目设置为 500。N-PCA 中使用最近邻分类器(NN)进行分类。具体实验结果如表 1 中所示，显然在更小的判别字典规模上，DKSVDN 的分类效果明显优于 LCKSVD 和 DLSI。此外，本文还比较了不同方法的时间效率，DKSVDN 预测类别所需时间最短。

表 1　AR 数据库人脸识别实验结果

Tab. 1　Experiment results on AR database

算法	精确率/%	CPU 耗时/s
LC-KSVD1	48.08	0.0011
LC-KSVD2	46.00	0.0009
DLSI	62.85	0.0029
FDDL	78.46	0.0035
FDDL-LCSRC	80.33	0.0043
RCSRC	66.89	0.0009
N-PCA(NN)	33.63	0.0133
OUR METHOD(预测编码)	80.12	0.0011
OUR METHOD(最优编码)	88.24	0.0016

在 Extend YaleB 人脸数据集上，DLSI 每个类别的类内字典项设置为 15，其他设置于文献[25]中一致。根据 Extend YaleB 数据库中的类别数，LCKSVD 字典数目设置为 496。N-PCA 中使用最近邻分类器(NN)进行分类。表 2 中展示了在 Extend YaleB 人脸数据集上的实验结果，在分类准确度和时间效率上依然是非常优秀的。

表 2　Extend YaleB 数据库人脸识别实验结果

Tab. 2　Experiment results on Extend yaleb database

算法	精确率/%	CPU 耗时/s
LC-KSVD1	56.12	0.0012
LC-KSVD2	88.24	0.0012
DLSI	89.45	0.0020
FDDL	93.75	0.0024
FDDL-LCSRC	91.45	0.0043
RCSRC	90.67	0.0010
N-PCA(NN)	60.28	0.0126
OUR METHOD(预测编码)	92.33	0.0012
OUR METHOD(最优编码)	97.24	0.0014

3.2.2 拟梦境训练实验

鉴于拟梦境训练方法针对于模式众多且分散的场景，本文选择在 AR 人脸数据库上验证拟梦境训练方法对 DKSVDN 的提升效果。AR 数据库包含正面人脸、佩戴墨镜、佩戴围巾三种模式，又因该数据集分为一期和二期两个子集，本文取一期中的一部分图片作为训练集。一期中包含每个人的前 13 张图片，即每个类别 13 张图片，其中正面人脸 7 张，佩戴墨镜 3 张，佩戴围巾 3 张，二期的图片构成与一期相同。将一期中的所有个体分为三个组别分别是：模式齐全组，缺失墨镜模式组，缺失围巾模式组。模式齐全组中的个体取所有图片，缺失墨镜组取正面人脸图像和佩戴围巾模式图像，缺失围巾组取正面人脸图像与佩戴墨镜模式图像。一期中的其他图像以及一期中的图像均用作测试集。为了更加直观地展示分组详情，将 100 个类别按 1 至 100 编号，训练集图像具体组成如表 3 所示。

表 3　训练集具体构成情况

Tab. 3　The details of the training set

类别	正面人脸	佩戴墨镜	佩戴围巾
1~40	包含	包含	包含
41~70	包含	包含	不包含
70~100	包含	不包含	包含

表 4 展示了预测神经网络模块在未进行拟梦境训练与进行拟梦境训练地 DKSVDN 模型在测试集上的识别效果。显然，拟梦境训练使得 DKSVDN 在识别率上有显著的提升，尤其是在墨镜模式和围巾模式两种遮挡模式下，识别率大幅优于未经拟梦境训练的 DKSVDN 模型。

表 4　拟梦境训练在 AR 数据库人脸识别实验结果

Tab. 4　Experiment results on AR database with Dreaming　/%

测试模式	正面人脸组		无围巾组		无墨镜组	
	拟梦境	原始模型	拟梦境	原始模型	拟梦境	原始模型
正面人脸	94.29	93.93	93.33	92.86	78.10	78.10
墨镜	89.17	87.50	94.44	93.33	83.89	80.56
围巾	85.83	85.00	59.44	58.89	50.00	47.78

4　结束语

可预测判别 KSVD 模型 DKSVDN 中字典结构由标签字典和描述字典组合而成，相对应地，基于该字典的稀疏编码向量分别包含了标签向量段与描述向量段，利用特殊结构的稀疏编码可以进行快速分类。同时，模型训练了一个预测神经网络模块用于预测稀疏编码，有效解决了传统字典学习时间效率低下的问题。此外，拟梦境训练方法有效提升了 DKSVDN 在样本多样性缺失的情况下的鲁棒性。本文在主流人脸数据集上的实验效果对比证明了 DKSDN 模型以及拟梦境训练方法在复杂环境具有优良的性能。

参考文献：

[1] 吴巾一，周德龙．人脸识别方法综述 [J]．计算机应用研究，2009, 26 (9): 3205-3209. (Wu Jinyi, Zhou Delong. Survey of face recongnition [J]. Application Research of Computers, 2009, 26 (9): 3205-3209.)

[2] Trigueros D S, Meng L, Hartnett M. Face recognition: From traditional to deep learning methods [J]. arXiv preprint arXiv: 181100116, 2018.

[3] Masi I, Wu Y, Hassner T, *et al.* Deep face recognition: A survey [C]// 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI) . IEEE: 471-478.

[4] 康利攀，陈方福，范自柱．一种新的拓展稀疏人脸识别算法 [J]．计算机应用研究，2016, 33 (3): 937-939. (Kang Lipan, Chen Fangfu, Fan Zizhu. New extended sparse representation for face recognition [J]. Application Research of Computers, 2016, 33 (3): 937-939.)

[5] Zhang Q, Li B. Discriminative K-SVD for dictionary learning in face recognition [C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 2691-2698.

[6] 沈韬，李克清，夏瑜．一种鲁棒稀疏表示的单样本人脸识别算法 [J]．计算机应用研究，2018, (11): 72. (Shen Tao, Li Keqing, Xia Yu. Single sample face recognition based on robust sparse representation [J]. Application Research of Computers, 2018, 35 (11): 3491-3496.)

[7] 赵雯，吴小俊．基于鉴别性低秩表示及字典学习的鲁棒人脸识别算法 [J]．计算机应用研究，2017, 34 (10): 3157-3161. (Zhao Wen, Wu Xiaojun. Robust face recognition of discriminative low-rank representation with dictionary learning [J]. Application Research of Computers, 2017, 34 (10): 3157-3161.)

[8] Lee K-C, Ho J, Kriegman D J. Acquiring linear subspaces for face recognition under variable lighting [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, 27: 684-698.

[9] Turk M, Pentland A. Face recognition using eigenfaces [C]// Proceedings 1991 IEEE computer society conference on computer vision and pattern recognition. 586-591.

[10] Wright J, Yang A Y, Ganesh A, *et al.* Robust face recognition via sparse representation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2009, 31: 210-227.

[11] Mi J-X, Sun Y, Lu J. Robust supervised sparse representation for face recognition [J]. Cognitive Systems Research, 2020.

[12] Neto J B C, Marana A N, Ferrari C, *et al.* Deep Learning from 3DLBP Descriptors for Depth Image Based Face Recognition [C]// 2019 International Conference on Biometrics (ICB) . IEEE, 2019: 1-7.

[13] Prasad P S, Pathak R, Gunjan V K, *et al.* Deep learning based representation for face recognition [M]// ICCCE 2019. Springer, Singapore, 2020: 419-424.

[14] Bashbaghi S, Granger E, Sabourin R, *et al.* Deep learning architectures for face recognition in video surveillance [M]// Deep learning in object detection and recognition. City: Springer, 2019: 133-154.

[15] Goswami G, Agarwal A, Ratha N, *et al.* Detecting and mitigating adversarial perturbations for robust face recognition [J]. International Journal of Computer Vision, 2019, 127 (6-7): 719-742.

[16] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[17] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[18] Baraniuk R G. Compressive sensing [lecture notes] [J]. IEEE signal processing magazine, 2007, 24: 118-121.

[19] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. Nature, 1996, 381: 607-609.

[20] Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: A strategy employed by v1? [J]. Vision research, 1997, 37: 3311-3325.

[21] Aharon M, Elad M, Bruckstein A. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation [J]. IEEE Trans on Signal Processing, 2006, 54: 4311-4322.

[22] Mairal J, Bach F, Ponce J, *et al.* Online dictionary learning for sparse coding [C]// Proceedings of the 26th annual international conference on machine learning. 2009: 689-696.

[23] 倪浩，阮若林，刘芳华．基于双正则化参数的在线字典学习超分辨率重建 [J]．计算机应用研究，2016, 33 (3): 277-281. (Ni Hao, Ruan Ruolin, Liu Fanghua. Image super-resolution based on online dictionary learning with two regularization parameters [J]. Application Research of Computers, 2016, 33 (3): 277-281.)

[24] Mairal J, Bach F, Ponce J. Task-driven dictionary learning [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 34 (4): 791-804.

[25] Jiang Z, Lin Z, Davis L S. Label consistent k-svd: Learning a discriminative dictionary for recognition [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35: 2651-2664.

[26] Mi J-X, Fu Q, Li W. Adaptive class preserving representation for image classification [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7427-7435.

[27] Ramirez I, Sprechmann P, Sapiro G. Classification and clustering via dictionary learning with structured incoherence and shared features [C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 3501-3508.

[28] Yang M, Zhang L, Feng X, *et al.* Fisher discrimination dictionary learning for sparse representation [C]// 2011 International Conference on Computer Vision. IEEE, 2011: 543-550.

[29] Wang D, Kong S. A classification-oriented dictionary learning model: Explicitly learning the particularity and commonality across categories [J]. Pattern Recognition, 2014, 47 (2): 885-898.

[30] Gregor K, LeCun Y. Learning fast approximations of sparse coding [C]// Proceedings of the 27th International Conference on International

Conference on Machine Learning. 399-406.

[31] Wang Z, Ling Q, Huang T S. Learning deep ℓ0 encoders [C]// Thirtieth AAAI Conference on Artificial Intelligence. 2016.

[32] Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition [C]// Proceedings of 27th Asilomar conference on signals, systems and computers. IEEE: 1993: 40-44.

[33] Dayan P, Hinton G E, Neal R M, *et al.* The helmholtz machine [J]. Neural computation, 1995, 7: 889-904.

[34] Hinton G E, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets [J]. Neural computation, 2006, 18 (7): 1527-1554.

[35] Martinez A M. The ar face database [J]. CVC Technical Report24, 1998.

[36] 王威，朱宗玖，陆俊. 基于字典学习和局部约束的稀疏表示人脸识别 [J]. 电脑知识与技术, 2018, 14 (5) . (Wang Wei, Zhu Zongjiu, Lu Jun. Sparse Representation for Face Recognition Based on Dictionary Learning and Locality Constraint [J]. Computer Knowledge and Technology, 2018, 14 (5): 3.)

[37] Mi J-X, Zhang Y-N, Lai Z, *et al.* Principal component analysis based on nuclear norm minimization [J]. Neural Networks, 2019, 118: 1-16.

[38] 米建勋，林志凯. 基于 L_2-范数重构样本约束的稀疏表示人脸识别方法 [J]. 计算机应用研究, 2020, 37 (4): 5. (Mi Jianxun, Lin Zhikai. Sparse representation classification via L2-norm based reconstruction sample constraint for face recognition [J]. Application Research of Computers, 2020, 37 (4): 5.)